

Building a Multi-Discipline Digital Library Through Extending the Dienst Protocol

Michael L. Nelson
NASA Langley Research Center
MS 157A
Hampton, VA 23607
m.l.nelson@larc.nasa.gov

Kurt Maly
Old Dominion University
Computer Science Department
Norfolk, VA 23529
maly@cs.odu.edu

Stewart N. T. Shen
Old Dominion University
Computer Science Department
Norfolk, VA 23529
shen@cs.odu.edu

1.0 Summary

The purpose of this project is to establish multi-discipline capability for a unified, canonical digital library service for scientific and technical information (STI). This is accomplished by extending the Dienst Protocol [1] to be aware of subject classification of a servers holdings. We propose a hierarchical, general, and extendible subject classification that can encapsulate existing classification systems.

2.0 Problem Statement

Spurred by recent advances in network information systems such as the World Wide Web, digital libraries are the topic of research in many scientific communities. However, digital library projects are partitioned by both the discipline they serve (computer science, aeronautics, physics, etc.) and by the format of their holdings (technical reports, video, software, etc.). There are over 10 existing or recent different digital library projects spanning over 5 different disciplines [2]. In short, each community is hand crafting their own digital library infrastructure.

One method for building an multi-discipline digital library is to extend an existing digital library protocol to be aware of subject classifications of both its holdings and the holdings of other same protocol servers. The resulting digital library toolkit could then be propagated to multiple disciplines and the various instantiations from multiple domains can register their existence with a meta-server.

3.0 Overview of Dienst

The Dienst protocol is the foundation for the National Computer Science Technical Report Library (NCSTRL) [3]. We choose Dienst because it is the most scalable, flexible, and extensible of digital library systems surveyed [2].

While Dienst is discipline independent, it is currently discipline monolithic. It makes no provision for knowledge of multiple subjects within its system. While it is possible to set up a collection of Dienst servers independent of NCSTRL, there is no provision for linking such collections of servers into a higher level meta-library.

Dienst consists of 5 components: 1) Repository Service; 2) Index Service; 3) Meta-Service; 4) User Interface Service; and 5) Library Management Service. Dienst names objects in collections using the CNRI Handle system [4]. Meta-data for objects is stored in the RFC-1807 format [5].

4.0 Approach

We propose a system where principals (typically authors) enter reports into a Dienst collection (such as NCSTRL) provide a subject classification for their entry. To minimize the impact on the contributing principals, preferences for the classification would be stored by the library management interfaces.

The yet to be specified classification system would be hierarchical and should be able to encapsulate any discipline-oriented classification system, such as the ACM Computing Review system or the NASA STI categories. Categories will be at a very high level and exist to partition various servers for the benefit of Dienst. Provisions will be made for appropriate cross listing of subjects, where appropriate.

It is important to note that while subject categories are introduced, repositories will continue to be centered around publishing organizations. While subject groupings facilitate human perusal, institution based grouping is much easier to manage. This is especially true for large organizations that publish in a variety of subject categories (e.g., DOE, DOD, and NASA laboratories).

To accommodate the subject classification system, additional fields will be added to 13 existing Dienst message types in 4 components, and 2 new message types created for Library Management component. Corresponding interface changes to the report entry phase and the user searching phase will allow for subject category selection, with minimal impact on either existing process.

5.0 Foundations for Future Work

Implementation of a subject classification system in Dienst establishes a basis for future improvements, including subject based browsing (e.g., "1996 Distributed Computing Reports, All Institutions"), and subject based update notification (e.g., e-mail notification of new reports on "Distributed Computing").

The results could be integrated with the Project Envision Software [6], which provides a rich user interface for navigating subject categories and search results. The updates could also be used for separate projects in multi-discipline research, exploring the similarities and differences between points of intersection of various disciplines.

6.0 Conclusions

This project proposes to extend the Dienst 4.0 protocol to incorporate subject classification. This is a necessary step toward building a canonical registry for digital libraries of multiple disciplines, utilizing the same protocol. The subject classification to be introduced will be high level and have the ability to encapsulate existing discipline specific categorization schemes.

7.0 References

1. Davis, J. R.; Krafft, D. B.; & Lagoze, C.; "Dienst: Building a Production Technical Report Server," *Advances in Digital Libraries*, Springer-Verlag, 1995, pp. 211-222.
2. Esler, S. L.; & Nelson, M. L.; "The Evolution of Scientific and Technical Information Distribution," Submitted to the *Journal of the American Society for Information Science*, 1997, <http://blurple.larc.nasa.gov/sti-evolution.doc>
3. The National Computer Science Technical Report Library, <http://www.ncstrl.org/>
4. Kahn, R. E.; "The Handle System Version 3.0: An Overview," <http://www.handle.net/docs/overview.html>
5. Lasher, R.; & Cohen, D.; "A Format for Bibliographic Records," Internet RFC-1807, June 1995.
6. Fox, E. A.; Heath, L. S.; & Hix, D.; "Project Envision Final Report: A User-Centered Database from the Computer Science Literature," July 1995, <http://ei.cs.vt.edu/papers/ENVreport/final.html>